Quantifying uncertainty of predictions from nonlinear cost estimation relationships (CERs)

October 11, 2007

Dr. Richard Sperling, CNA Corporation

CERs and prediction uncertainty

- A CER models cost of a weapon system as a function of system characteristics
  - Example: $C = \alpha + \beta_1 Weight + \beta_2 Speed$
- CERs are also used to predict the cost of weapons systems
- Typically, CERs are estimated with limited data—10 to 15 observations are not uncommon
- Smaller data samples lead to greater uncertainty in cost predictions

Two sources of CER uncertainty

- Sampling or chance error
  - We are estimating CER parameters from a finite sample of data
- Specification error
  - CERs are parsimonious models of the system whose cost we are predicting
  - We cannot account for every factor that affects system cost

Measure of prediction uncertainty for linear and non-linear CERs

- We use the standard error of the cost prediction to measure the uncertainty of CER predictions
- There is an exact formula for the standard error of a cost prediction from a linear CER
- But some CERs are nonlinear
- There is no exact formula for the standard error of a cost prediction from a nonlinear CER, so we need a method to approximate the standard error
- More generally, there are no exact results for non-linear models, only approximations
  - We appeal to asymptotic (large-sample) statistical theory

Choice of non-linear model

- We want a model that is asymptotically identified
  - As the sample size gets larger, the model is always identified
  - The parameters are estimable
- Not all non-linear CERs are asymptotically identified
- What does a model that is or is not asymptotically identified look like?

Illustration of a model that is not asymptotically identified

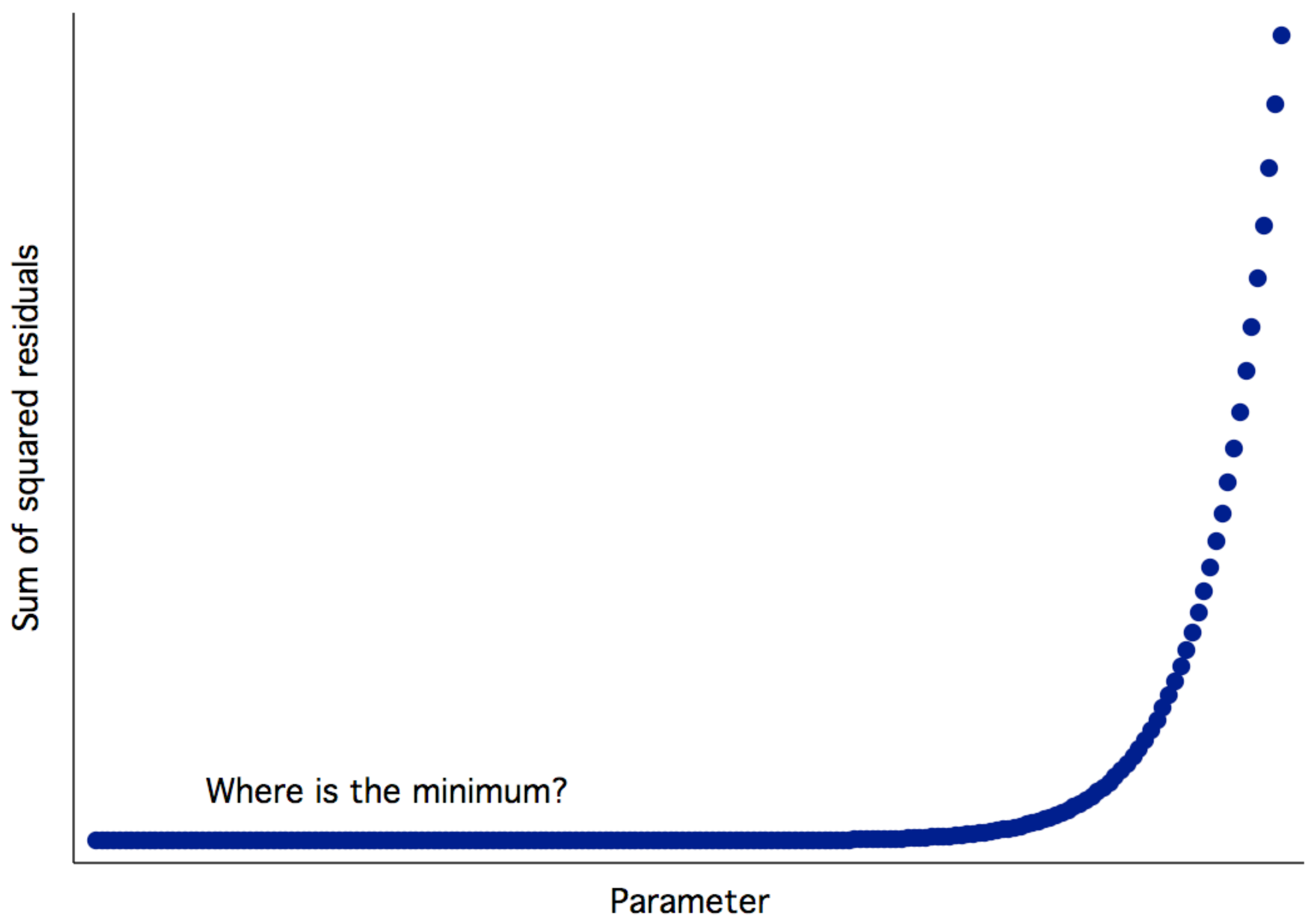
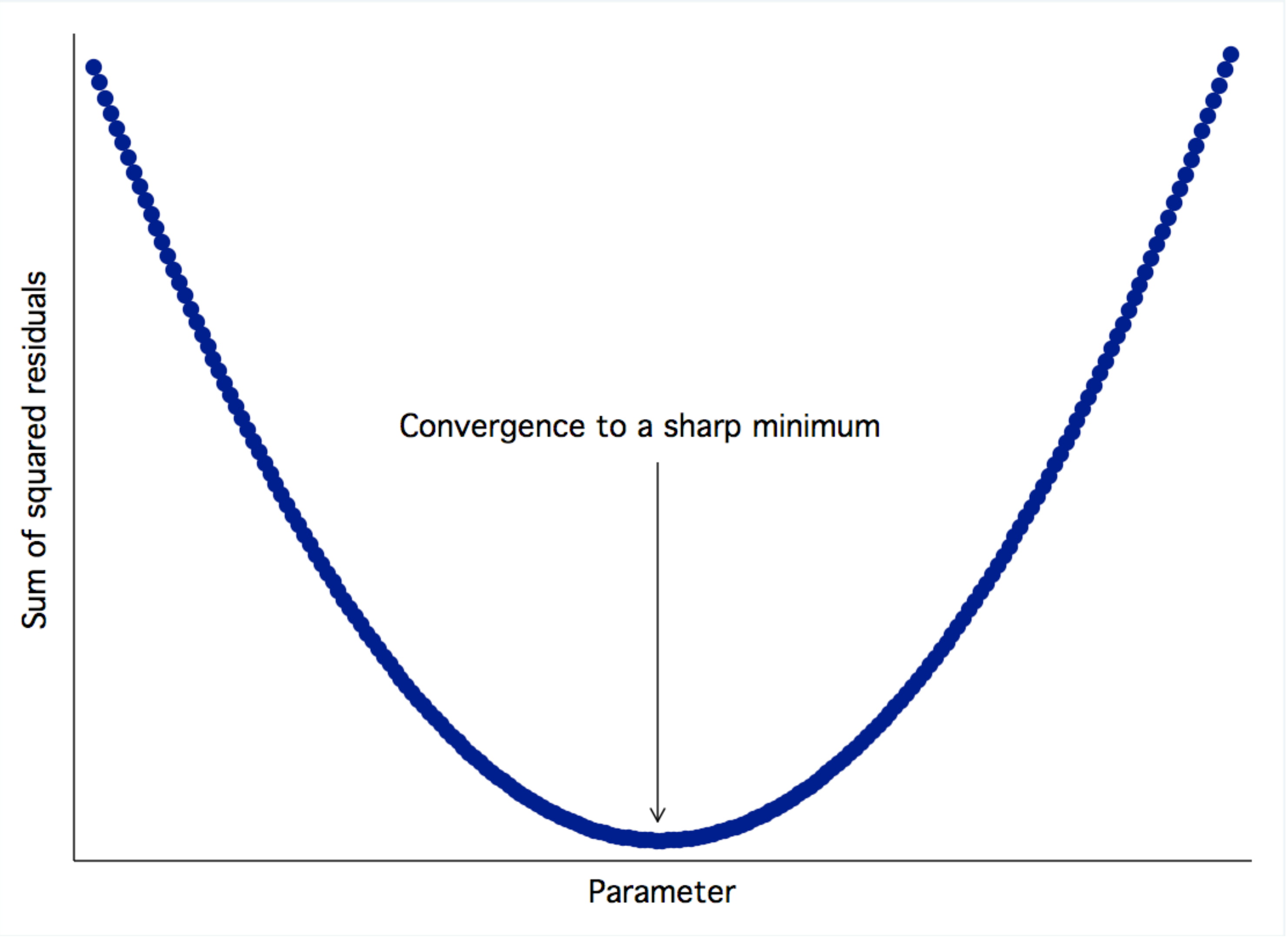


Illustration of an asymptotically identified model



General non-linear CER

- Multiplicative error CER:

$$C_t = f(\mathbf{x}_t;\boldsymbol{\beta}) \times \epsilon_t \quad \epsilon_t \sim \mathrm{IID}(1,\sigma^2),$$
$$t = 1,\ldots,n$$

- This functional form implies:

$$\begin{aligned}\mathrm{Var}(C_t) &= [f(\mathbf{x}_t;\boldsymbol{\beta})]^2 \times \mathrm{Var}(\epsilon_t) \\ &= [f(\mathbf{x}_t;\boldsymbol{\beta})]^2 \times \sigma^2\end{aligned}$$

In words, the variance of cost is proportional to the square of the cost prediction ($f(\mathbf{x}_t;\boldsymbol{\beta})^2$), i.e., it is more difficult to predict cost for higher than lower cost systems

Choice of non-linear estimation method to estimate β

- We want a consistent statistical estimator
- Not all non-linear estimators are consistent
- An estimator is consistent only if the non-linear model is asymptotically identified
- What is consistency?

Consistency

- Is another asymptotic (large sample) result
- Simple explanation: you are estimating the right thing
- As the sample size gets larger, the statistical estimator (or formula) yields a parameter estimate that becomes closer to the true, but unknown, value of the population statistic that we are estimating
- If the statistical estimator is not consistent, then, no matter how large the sample, the estimate will never converge to the true, but unknown, value of the population statistic

Estimation methods

- We consider two methods for estimating parameters from a non-linear CER with a multiplicative error
- Iteratively re-weighted least squares or **IRLS** (weighted non-linear least squares in this case) **is consistent**
- **MPE-ZPB** (minimum percentage error, zero percentage bias) **is not consistent**

IRLS vice MPE-ZPB

- IRLS is a quasi-maximum likelihood estimator*
  - It is consistent under minimal assumptions
  - It may be biased in small samples
  - Has a known formula for the variance-covariance matrix
  - The IRLS estimator is asymptotically normally distributed even if the regression error is not
- MPE-ZPB is biased and inconsistent
  - ZPB restricts the sum of the residuals (estimated errors) to zero
  - But zero residuals in a sample do not guarantee unbiased or consistent parameter estimates
  - Variance-covariance matrix is unknown

*Reference: Matthew Goldberg and Anduin Touw, *Statistical Methods for Learning Curves and Cost Analysis.* INFORMS, Topics in Operations Research Series, 2003.

From cost estimation to cost uncertainty for non-linear CERs

- There are no exact statistical results when working with non-linear CERs
- We have examined desirable properties of non-linear CERs and methods to estimate their parameters
- We should choose an estimator based on its statistical properties
- We can use our non-linear CER to predict cost after we have estimated the parameter vector $\boldsymbol{\beta}$
- Given predicted costs, how do we quantify the uncertainty of predictions from a non-linear CER?

Methods for quantifying the uncertainty of predictions from non-linear CERs

- Delta method
- Non-parametric Bootstrap
- The Delta method and Bootstrap are used when you want to approximate a statistic (e.g., variance or standard error of the cost prediction) that is hard to calculate or for which no formula exists

Delta method

- The Delta method is a one-time non-linear calculation
  - It’s fast
  - It only requires a single computational pass through the data
  - Known since at least Tukey (1957) and Rao (1965)
- The Delta method uses a truncated (usually first-order or linear) Taylor series expansion around a point (e.g., the mean of the data) to approximate the variance of a statistical estimator
- The approximation is good in a neighborhood around the point of expansion
  - The approximation worsens the farther you move from the expansion point
- Why a Taylor Series expansion? Calculation of functions is sometimes impossible or difficult; a Taylor series uses polynomial functions to approximate more complicated functions
- The Delta method is implemented in the statistical program *Stata*

Bootstrap

- The Bootstrap is a statistical re-sampling method for estimating the bias, variance, or standard error of a statistical estimator or statistic
  - A variant of Monte Carlo simulation
  - Developed in the 1970s by Dr. Bradley Efron
  - Computationally intensive, but less of an issue now given today’s powerful computers
- The Bootstrap re-samples with replacement from the residuals (estimated errors)
  - The underlying assumption is that estimated residuals accurately represent the true, but unknown, error distribution
- The Bootstrap is particularly useful in cases where asymptotic calculations are difficult or unknown
- Davison and Hinkley (1997) provide a good book-length exposition
- The Bootstrap is implemented in the statistical program *Stata*

Comparing the Delta method to the Bootstrap in a specific case

- We ran a Monte Carlo experiment to compare the relative performance of the Delta method and Bootstrap in approximating the standard error of the prediction from a non-linear CER of the "triad" form

$$C_t = (\alpha + \beta X_t^{\gamma}) \times \epsilon_t, \quad \epsilon_t \sim \mathrm{NID}(1, \sigma^2)$$

- We estimated α, β, and γ using IRLS

Problem with the "triad" CER

- The "triad" CER is poorly identified*
- More specifically, the data are highly correlated which makes it hard to estimate the CER parameters with reasonable precision
- If the data are highly correlated, it is difficult to "disentangle" or estimate the separate effects of α, β, and γ

*Reference: Russell Davidson and James G. MacKinnon, *Estimation and Inference in Econometrics*. Oxford University Press, New York, 1993.

General results from the Monte Carlo experiment

- The Delta method was fast and provided reasonable estimates of the standard error of the CER prediction for all but the smallest sample sizes
- The Bootstrap was very time-intensive and yielded no results due to poor identification of the “triad” CER
  - This result does not mean that the Bootstrap is not useful generally, only that it worked poorly in the case of using IRLS to estimate the “triad” CER parameters

Conclusions

- CER model
  - In general, avoid using the “triad” CER with a multiplicative error because it is poorly identified
- Statistical estimator
  - IRLS is consistent, whereas MPE-ZPB is not
- Quantifying uncertainty of cost predictions from non-linear CERs
  - The Delta method provides fast and reasonable estimates of CER prediction uncertainty in many cases, but the approximation becomes worse as you move farther away from the expansion point
  - The Bootstrap is computationally intensive, but is useful particularly in cases where a CER is highly non-linear

References

- Russell Davidson and James G. MacKinnon. *Estimation and Inference in Econometrics*. Oxford University Press, New York, 1993.
- A.C. Davison and D.V. Hinkley. *Bootstrap Methods and their Application.* Cambridge University Press, New York, 1997.
- Matthew S. Goldberg and Anduin E. Touw. *Statistical Methods for Learning Curves and Cost Analysis.* Topics in Operations Research Series. INFORMS, Linthicum, MD, 2003.
- C.R. Rao. *Linear Statistical Inference and Its Applications.* John Wiley, New York, 1965.
- John W. Tukey. The Propogation of Errors, Fluctuations and Tolerances: Basic Generalized Formulas. Technical Report 10, Princeton University, Statistics Research Group, December 1957.

Backup

Illustration of the Delta method computation for the general case

- General nonlinear differe $F(Y_1 \cdots Y_m)$ function:
- Gradient vector:

$$\mathbf{g} = (\partial F / \partial Y_1 \cdots \partial F / \partial Y_m)$$

- Delta method:

$$\mathrm{Var}[F(Y_1, \ldots, Y_m)] \approx \mathbf{g}\, \mathrm{Cov}(\mathbf{Y})\, \mathbf{g}^T$$

Illustration of the Delta method computation for the "triad" CER

- Functional fo $F(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) = \widehat{C} = \hat{\alpha} + \hat{\beta} x^{\hat{\gamma}}$
- Gradient vect $\mathbf{g}^T = \begin{pmatrix} \partial \widehat{C} / \partial \hat{\alpha} \\ \partial \widehat{C} / \partial \hat{\beta} \\ \partial \widehat{C} / \partial \hat{\gamma} \end{pmatrix} = \begin{pmatrix} 1 \\ x^{\hat{\gamma}} \\ \hat{\beta} x^{\hat{\gamma}} \ln(x) \end{pmatrix}$
- Delta metho $\mathrm{Var}(\widehat{C}) \approx \mathbf{g} \, \mathrm{Cov}(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) \, \mathbf{g}^T$

$$= \begin{pmatrix} 1 & X^{\hat{\gamma}} & \hat{\beta} X^{\hat{\gamma}} \ln(X) \end{pmatrix} \left[\hat{\sigma}^2 \times \mathbf{V} \right] \begin{pmatrix} 1 \\ X^{\hat{\gamma}} \\ \hat{\beta} X^{\hat{\gamma}} \ln(X) \end{pmatrix}$$

The goal of a Monte Carlo experiment

- We want to compare the statistical properties of the Bootstrap and Delta method as methods for quantifying the uncertainty of cost predictions from a non-linear CER
- We make these comparisons using the sampling distributions of the two methods or estimators
  - A sampling distribution is a probability (frequency) distribution of an estimator for an infinite number of samples
  - We can’t calculate results for an infinite number of samples, so we run a “large” number of Monte Carlo replications instead
- Two properties of sampling distributions that are of great interest are the mean (central tendency) and variance or standard deviation (dispersion)
- We derive sampling distributions theoretically or experimentally
- Monte Carlo simulation is an experimental method to derive sampling distributions of estimators

Monte Carlo Experiment

- Generate data sample from an assumed known, or true CER of the "triad" form
- For Monte Carlo replications 1 to 1,000:
    1. Estimate the CER parameters by Iteratively Re-weighted Least Squares
    2. Predict the cost of a new system
    3. Tabulate the predicted cost
    4. Estimate standard deviation of prediction using the Delta method
    5. Estimate standard deviation of prediction using Bootstrap
- Calculate average and standard deviation of 1,000 values of predicted cost
- Calculate average of 1,000 Delta method standard errors
- Calculate average of 1,000 Bootstrap standard errors

Parameters of Monte Carlo experiment

- Assumptions:
  - True CER of the "triad" $C_t = \alpha + \beta x_t^\gamma$
  - CER with multiplicative $C_t = (\alpha + \beta x_t^\gamma) \times \epsilon_t$
  - True parameter values:
    $\alpha = 1, \beta = 2, \gamma = 0.4, \quad \epsilon_t \sim \mathrm{NID}(1, 0.15^2)$
  - Assume cost driver (*x*) is uniformly distributed in [10,100]
  - Prediction for new system at *x** = 80
- Repeat experiment for sample sizes (*n*) of 10, 25, 50, and 100

Results of Monte Carlo experiment

1,000 Monte Carlo replications		Sample size (n)			
		10	25	50	100
True value across Monte Carlo replications	SE of cost prediction	(a)	0.5278	0.3690	0.2810
Delta method	Avg value, SE of cost prediction at x=80	(a)	0.4907	0.3433	0.2642
	Mean of cost driver(x)	68.3	62.0	60.5	56.1
Bootstrap	Avg value, SE of cost prediction	(a,b)	(b)	(b)	(b)

(a) Monte Carlo simulation took about 5 hours and only completed 250 replications

(b) Unable to compute results due to high computational cost of the Bootstrap in this case

Bootstrap computational cost

135 Bootstrap replications (IRLS estimation of "triad" CER)	Sample size (n)			
	10	25	50	100
SE of prediction	0.3785	0.5888	0.3016	0.2669
Computational time	6 h 10 m	2 h 20 m	1 h 10 m	6 m

Note: Computations performed using version 9.2 of Stata MP on a MacBook Pro 2.33 GHz Intel Core2 Duo with 2 GB of 667 MHz DDR2 SDRAM

- Assume that the average time to complete 135 Bootstrap replications of the IRLS estimator applied to the "triad" CER is 6 minutes
- A Monte Carlo experiment with 1,000 replications that simulates this Bootstrap procedure (135,000 IRLS estimations) would take about 6,000 minutes or 100 hours (a little more than 4 days)